

# The Eventual Decline of Empirical Law and Economics

Saul Levmore<sup>†</sup>

*This Essay suggests the necessity of a co-evolutionary process among empirical and theoretical advances in law and economics. Empirical work alone is suggestive, but should not be taken too seriously. The weaknesses in empirical work, and by this I mostly mean regression-based work which has come to dominate law and economics, lead to a kind of virus that begins with over-statements and misapprehensions, and then spreads as more scholars copy the mistakes and engage in empirical work as a means of entry into the field. Regression-based work will become suspect as its current assumptions are questioned, and as replication failures reveal its weaknesses. Empirical work in law and economics looks very different when underlying distributions are not easily probed with regressions but are understood as reflecting power-laws, or as simply random. Once inconvenient distributions are acknowledged, the key question is why observations might be distributed in this fashion. This is likely to be a task for theorists as law and economics enters its next phase. On the other hand, empirical work has been important and has made law and economics a respectable science. The claim here is that good empirical work—especially in law and economics—is hard to produce, and it is important not to over-value its products. Moreover, it is more useful when combined with good theory.*

*The focus in this Essay is on three weaknesses of empirical work, though in a larger sense most of the problems come from omitted variables and, in some cases, insufficiently large data sets. First, much of the empirical work in law and economics is driven by models that rely on error minimization techniques, and these techniques are unreliable when errors are surprisingly and unevenly distributed (that is, when they suffer from heteroscedasticity). Second, it is likely that when empiricists connect data with a model, the process is flawed because there might be a hidden transition to a second distribution. Discovering multiple distributions is likely to require theoretical work. These and other problems are exacerbated by the likelihood that conclusions are based on the tail end of data sets, inasmuch as scholarly journals only bring to light statistically significant results. In*

---

<sup>†</sup> William B. Graham Distinguished Service Professor of Law, University of Chicago. I am grateful for suggestions I received from Frank Fagan, Julian Gale, Dorothy Lund, Eric Posner, and workshop participants at the University of Chicago Law School and at the conference that motivated this work. I am also grateful for the support received from the Lawrence T. Hoyle, Jr. Faculty Fund.

## The Eventual Decline of Empirical Law and Economics

*addition, empirical work in law and economics suffers from the absence of sizeable data sets. Without such sets it is difficult to test conclusions and to escape the omnipresent challenge of omitted variables. Reversal paradoxes are yet another serious problem, and especially so in the absence of large data sets.*

*The larger and more optimistic claim is that data and theory can and must work together. Regressions have come to play a critical role in law and economics, and econometric methods have improved over time. It has become apparent that data can suggest theories, and theories can be tested, to a degree, with data. But some theoretical insights are so convincing that data testing, though comforting even when flawed, may be unnecessary—and it may, in any event, be tainted by the spread of the theory. It is likely that empirical work in law and economics will find itself in retreat, even as its quality improves because of renewed attention to theory.*

Introduction .....	613
I. Questioning Empirical Work: Power Laws and Other Inconvenient Distributions .....	616
II. The Mistaken Assumption of Single Distributions .....	621
III. Limited Data and Reversals.....	624
IV. Theory-Data Interdependence .....	626
Conclusion.....	631

### Introduction

This Essay will not be the first to emphasize the weakness of empirical work that is not motivated by theory.<sup>1</sup> My larger aim is to stress and predict the continuing co-evolution of these two strands of law and economics.<sup>2</sup> Along the way, I develop several themes. First, and perhaps most controversial, is the idea that a great deal of empirical work in law and economics (as in other fields) is suggestive, but is not to be taken too seriously. Indeed, I predict that the replication crisis, best associated with psychology, will soon find its way to law and economics.<sup>3</sup> The weaknesses in empirical work,

---

1. See, e.g., JUDEA PEARL & DANA MACKENZIE, *THE BOOK OF WHY: THE NEW SCIENCE OF CAUSE AND EFFECT* (2018); Richard N. Boyd, *On the Current Status of the Issue of Scientific Realism*, 19 *ERKENNTNIS* 45 (1983); Tjalling C. Koopmans, *Measurement Without Theory*, 29 *REV. ECON. & STAT.* 161 (1947). By “theory” I mean something interesting that explains how a part of the world works.

2. For a general discussion on the early development of empirical legal scholarship, see Michael Heise, *The Past, Present, and Future of Empirical Legal Scholarship: Judicial Decision Making and the New Empiricism*, 2002 *U. ILL. L. REV.* 820.

3. See Scott E. Maxwell, Michael Y. Lau & George S. Howard, *Is Psychology Suffering from a Replication Crisis? What Does “Failure to Replicate” Really Mean?*, 70 *AM. PSYCHOLOGIST* 487 (2015). Replication crises are not limited to psychology. There is good reason to think that most of the social sciences and many of the hard sciences are replete with

by which I usually mean regression-based work that is dominant in law and economics, are a kind of virus that begins with over-statements and misapprehensions and then spreads as scholars copy the mistakes of their predecessors, and construct more regressions as a means of entry into the field. Good empirical work is simply hard to do, and we need to be aware of its limitations. I focus on three weaknesses of empirical work in a field that rarely has the luxury of large data sets. First, much of the empirical work in law and economics is driven by models that are formed by error minimization techniques, and these techniques are unreliable when errors are surprisingly and unevenly distributed (that is, when they suffer from heteroscedasticity), as is likely to be the case when the true distribution of variables is not normal, but instead conforms to a power law, to one of various sigmoid (convex and then concave, s-curve) functions, or reflects midstream changes from one distribution to another, including a segment of randomness. I aim to show that while empiricists like to say that “some data is better than no data,” the seemingly obvious reliance or insistence on data-driven analysis, is misplaced or even reckless, because there is often reason to think that “some data” often leads to misleading conclusions. As we will see, omitted variables are at the root of the problem. Most empiricists are of aware of this problem, but the problem is insufficiently appreciated, and the responses to it come with problems of their own.

The weaknesses commonly encountered in empirical work reflect the well-known difficulty of moving from correlations to claims of causation. Great strides have been made in recent times, in what is known as the “causation revolution,” to bridge this gap,<sup>4</sup> but much of the work has relied upon, or at least been made easier with, linear models. Linearity is often a convenient rather than a supportable assumption, as explained in Part I. At times the assumption of linearity does little harm because there are natural or clever experiments available, random sampling may be possible, and enough data is available to test predictions. This is done by setting aside some unseen data, or at least dividing it by predicting future developments, in ways that reduce the likelihood that omitted variables do the work.<sup>5</sup> Testing regressions on set aside data can be seen as a kind of internal replication. Ideally, and especially in law and economics, insights are based on theories. Theories about *why* correlations are found lead to predictions and tests. Put differently, econometrics is a game of looking for treatment effects, but doing so requires that the groups being compared are truly comparable, so that the groups are exchangeable and attention can be focused on the variable being studied. In practice, there is a

---

unreproducible, inconsistent, published, and oft-cited, studies. See Alvaro de Menard, *What's Wrong with Social Science and How to Fix It: Reflections After Reading 2578 Papers*, FANTASTIC ANACHRONISM (Sep. 11, 2020), <https://fantasticanachronism.com/2020/09/11/whats-wrong-with-social-science-and-how-to-fix-it> [<https://perma.cc/F9VA-TKJR>].

4. PEARL & MACKENZIE, *supra* note 1, at 9.

5. See *id.*; Frank Fagan & Saul Levmore, *The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion*, 93 S. CAL. L. REV. 1, 16 (2019).

persistent problem of pollution by omitted variables. In some fields, randomized experiments with a large number of observations solve this problem. This is normally impossible in law and economics.

Part II then suggests a second problem that persists even if we can overcome the inconvenience of non-linearity. It must often be the case that what appears as one phenomenon, with empiricists striving to describe it with one line or curve, is really best described with two or more distributions—and perhaps theories. Readers might think of this as a subset, or extreme version, of the problem of omitted variables, the mainstay of the discussion in Part I, but it is useful to think of it as a separate matter. In passing, the discussion highlights yet another problem that is related to the familiar complaint that scholarly journals only bring to light statistically significant results; we may be looking at a tail end of data in the first place. Part III sets aside reasons to be nervous about imagined distributions and focuses on one of the many reasons to question conclusions drawn from relatively small data sets, as is typical in law and economics. A simple example suggests why this problem of Simpson’s Paradoxes, or reversals, is beguiling but also serious in the absence of a very large number of observations.

Part IV turns more directly to the partnership between data analysis and theory. The discussion considers problems for which data suggests theories and, in contrast, where theories come first, and are then tested with data. Some theoretical insights are so convincing that data testing, comforting even though flawed, may be unnecessary—and testing may, in any event, be tainted because the decisionmakers who are observed are aware of the theoretical insights, and able to respond to them. The discussion concludes that we should expect empirical and theoretical work to evolve in combination. The recent surge of empirical work in law and economics will be slowed not only by the problems described in Parts I, II, and III, but also by the exhaustion of theories that have been evaluated. Further empirical work will eventually require the development of new theories.

The analysis here repeatedly refers to the relative paucity of data available to empiricists in law and economics. Statistical techniques often involve sampling, whether intentional or forced, which is to say taking a set of observations in order to say something about the larger pool of actual events or potentially available observations.<sup>6</sup> At times, the larger set will be experienced in the future, while in other settings the smaller set is used to predict the larger one because it is costly to gather data.<sup>7</sup> In both cases

---

6. Ever since William Gossett wanted to test the quality of the hops being used by Guinness, sampling and measuring errors has been the main source of employment for many a statistician. See Dan Kopf, *The Guinness Brewer Who Revolutionized Statistics*, PRICEONOMICS (Dec. 11, 2015), <https://priceonomics.com/the-guinness-brewer-who-revolutionized-statistics> [<https://perma.cc/4KDW-X28H>].

7. PEARL & MACKENZIE, *supra* note 1, at 155, offer an excellent if counterintuitive explanation, when discussing the “potential outcomes” framework deployed by modern

there is the problem highlighted in Part I; the techniques that extrapolate from a small set to a larger one make assumptions that are often false, and trouble follows. The distribution of real-life data (and the errors thus identified in the subset) may be inconvenient for the empiricist. A good theory can avoid this problem, but this requires data analysis to be guided, or at least instigated, by a theory. It argues for moving from theory to data, rather than from data to theory (or no theory). Empirical work often tests theories by testing over time, as time offers a means of setting aside data. Unfortunately, this strategy brings on a new set of omitted variables, simply because things change over time. Straightforward examples are provided to support these arguments, and they suggest that data science, at least with respect to law and economics, is likely to experience its own s-curve—a rise and then a fall, or a leveling off.

### I. Questioning Empirical Work: Power Laws and Other Inconvenient Distributions

It is plain that law and economics has left its first theoretical phase and prioritized empirical work. The leading journals are now filled with empirical work, and newcomers to the academy are likely to be trained in empirical methods. With enticing claims about what their empirical work suggests or proves, student-run journals have become useful partners in the rush to overclaim and gain attention. In some fields, like corporate law, international law, and criminal law, entry at the top is virtually impossible

---

econometricians. In testing whether a flu vaccine actually *causes* flu (as some critics suggest, because they observe that some people who take the vaccine do get the flu), they point to four possibilities, or groups, and call them: doomed, causative, preventive, and immune. *Doomed* means they get the flu even if vaccinated. *Causative* means they get the flu only if they are vaccinated (as alleged and now being tested). *Preventive* means the vaccination works with some significant probability. *Immune* means they do not get the flu whether or not vaccinated. The genius is that there is no need to worry about (and control for) every conceivable confounding variable. But the critical assumption of this strategy is that the four groups are evenly balanced in the data, or divided in a known configuration. Without this balance or knowledge, we cannot say whether the treatment and control groups are exchangeable. The very point is that counterfactuals are unobservable, so that the groups are surely unbalanced. Econometricians are left comparing apples and oranges.

This example skirts around the additional problem created by *unimagined* omitted variables. The vaccine may be dangerous for someone who consumes alcohol, or perhaps entirely effective for one who does not. In the language of the potential outcomes approach, there may be other paths from  $A$  to  $X$ . When omitted variables can be identified, they can be tested in order to say something about causation between  $A$  and  $X$ . But if the omitted variable is unanticipated, then some theory is required to be comfortable with the idea of looming omitted variables. Imagine again that we are testing whether  $A$  causes  $X$ , after observing a correlation between the two.  $B$  and  $C$  might be relevant omitted variables, and we aim to find out whether  $B$  and  $C$  provide another path between  $A$  and  $X$ , so that the  $AX$  correlation does not mean that one causes the other. We might find that  $A$  is attached to  $B$  or that  $A$  is attached to  $BC$  when those two are both found. We can now see whether  $A$  and  $X$  connect in the absence of  $B$  and  $C$ . This is the genius of the potential outcomes approach, but it requires some testing for the presence and absence of  $BC$ .  $AX$  may in fact be connected through  $EF$ , and if we are unaware of that possible connection, we cannot test for it. It may be  $E$  that brings on  $X$ .

## The Eventual Decline of Empirical Law and Economics

without empirical training and ambitions. Special attention is given to people who come equipped with good data sets, even when it is hard to say anything definitive or novel with these sets and no new theory.<sup>8</sup> This is the age of empirical law and economics. In the beginning of this phase, correlations were revealed. To take a well-known example, a heavily examined question in corporate law was whether the popularity of Delaware as a state of incorporation was a “race to the bottom” or, through familiar competitive pressures, to the top.<sup>9</sup> The empirical work began with two competing theories, and sought to declare one the winner. Early work revealed that firms that moved to Delaware were more likely to experience a *rise* in market value rather than a drop, so that Delaware appeared to offer good news for shareholders, rather than a race by migrating corporations to the bottom—that is, to permissive state laws that enticed managers at the expense of shareholders, who presumably needed protection by federal law or revisions in Delaware law.<sup>10</sup> This was a victory for a serious theory, in the sense that it was a view contrary to conventional wisdom.

However, as every statistics student learns, correlation is not causation. Perhaps there was indeed a race to the bottom, but migration to Delaware signaled a forthcoming corporate takeover that would benefit shareholders and overwhelm any decrease to be associated with management-friendly Delaware law. The migrating firms might have done yet better in the stock market if they had remained in their pre-Delaware states. It would have been nice to have counterfactuals. Perhaps out of every 100 firms on their way to Delaware, there was some way to bar a significant number of randomly chosen firms from relocating, and then to compare the change in their market prices with those that relocated. Alternatively, perhaps there was a way to isolate firms that relocated to Delaware and did not merge—though it would be hard to know whether new and old shareholders (incorrectly) anticipated a merger or other significant transaction. A generation or two later, the literature is still full of empirical work on these causation questions. Over time it became generally accepted that Delaware was good for shareholders, though there are enough confounding variables—including the availability of other states of incorporation, the expected decisions of Delaware courts, and the intervention of federal law—to keep corporate law scholars busy, and their readers exhausted.<sup>11</sup>

---

8. For one take on this, see Tom Ginsburg & Thomas J. Miles, *Empiricism and the Rising Incidence of Coauthorship in Law*, 2011 U. ILL. L. REV. 1785. There is likely to be some push-back based on the likelihood that a focus on impressive data sets might be unfair because well-connected young scholars have better access to data sets.

9. Daniel R. Fischel, *The “Race to the Bottom” Revisited: Reflections on Recent Developments in Delaware’s Corporation Law*, 76 NW. U. L. REV. 913 (1982).

10. For an explanation of these competing theories and a review of the associated early empirical work, see Jonathan R. Macey & Geoffrey P. Miller, *Toward an Interest-Group Theory of Delaware Corporate Law*, 65 TEX. L. REV. 469, 469 nn.1-2 (1987).

11. Early work, and the entire field, is best credited to Roberta Romano, but useful attractions include William J. Carney & George B. Shepherd, *The Mystery of Delaware Law’s*

Most empirical work in law and economics, and especially work that tries to validate or invalidate claimed causal connections (whether a move to Delaware will be good for shareholders, for instance), is regression analysis, a form of hypothesis testing (or rejection). Consider, for example, a store (or army or prison) that wants to decide what size men's clothes to stock. Over many years, the height of men has been recorded as they enter military service, and we have learned that there is a normal (Bell curve) distribution of heights around a (current) median of about 5'10". This classic example of a normal distribution means that the 25<sup>th</sup> and 75<sup>th</sup> percentiles (2 inches shorter and taller as it turns out) have the same number of people, and the 10<sup>th</sup> and 90<sup>th</sup> percentiles (4 inches) also match one another in size.<sup>12</sup> The same neat distribution holds for women, for people at different ages of life, in different countries, and in different years, though the median changes (it was 2 inches lower in the United States one hundred years ago, and it is now higher in the Netherlands, and so forth). Our buyer might need to adjust for the income of people expected to wear these uniforms, but that too is easily done. If the buyer needed to acquire clothes to be used many years in the future, the task gets harder, because the median height has increased over time, and may continue to do so as recent immigrants are encompassed. Increases in the median are usually thought to be the result of diet and other factors,<sup>13</sup> and while it went through periods in which the rate of this increase seemed regular or predictable, it now appears to have leveled, and certainly to have slowed. This s-curve phenomenon (predictable and accelerated increase followed by a slowing increase and then a leveling) is true of so many things that long-term predictions are nearly impossible without some theory about the limits of growth.<sup>14</sup> There were, by way of other examples, a rapid, exponential increase in the demand for horses, and then for trains and for cars and eventually for aircraft, but in the first of these examples demand eventually decreased rather steeply, and in the third example it has leveled off. Note that some of these decreases, like empirical work itself, were accompanied by decreasing costs. Bicycles and regressions might be cheaper to produce over time, and yet they can decrease in importance because demand changes, as substitutes

---

*Continuing Success*, 2009 U. ILL. L. REV. 1; Joseph A. Grundfest, *The Limits of Delaware Corporate Law: Internal Affairs, Federal Forum Provisions, and Sciabacucchi*, 75 BUS. LAW. 1319 (2020); Marcel Kahan & Edward Rock, *Symbiotic Federalism and the Structure of Corporate Law*, 58 VAND. L. REV. 1573 (2005); Robert B. Thompson, *Delaware's Disclosure: Moving the Line of Federal-State Corporate Regulation*, 2009 U. ILL. L. REV. 167; Lucian Arye Bebchuk & Assaf Hamdani, *Vigorous Race or Leisurely Walk: Reconsidering the Competition over Corporate Charters*, 112 YALE L.J. 553 (2002).

12. See Max Roser, Cameron Appel & Hannah Ritchie, *Human Height*, OUR WORLD IN DATA (May 2019), <https://ourworldindata.org/human-height> [<https://perma.cc/KZY9-8DXN>].

13. The language in the text is a bit imprecise in order to save space. The underlying point is that we usually have evidence of correlation, and there is then some guesswork about causation.

14. See VACLAV SMIL, *GROWTH: FROM MICROORGANISMS TO MEGACITIES* (2019).

become available or buyers learn that bicycles (or regression results) are less useful than first imagined. S-curves can often be explained after the fact, but exceptions abound. Few of us would bet that climate warming will level off or decrease simply because the use of fuels that are currently popular is likely to level off.

Regression techniques are at their best when the real world that lies “beneath” studied observations follows a linear pattern—either because it is strictly linear or because it is exponential and can therefore be turned into a linear pattern through a logarithmic transformation.<sup>15</sup> For good reason, the regression line will be created with a technique that minimizes least squares (of the errors, or observations, that do not exactly fit the line revealed by the data).<sup>16</sup>

The major problem emphasized here is that common regression techniques work best with data that fits a straight line of the kind just described, but requires serious adjustment if the underlying distribution is non-linear. I will emphasize, or simply toss out in the interest of conserving words, two variants from any convenient distribution, such as a straight line or a normal (Gaussian) distribution. First, actual distributions may follow power (or scaling) laws, so that the tail (or even both tails) of a distribution is much steeper than that found (or expected even) in an exponential growth model. Power-law distributions are often traced back to Pareto’s observation that 80% of Italian land was owned by 20% of its people, and then this 80/20 “rule” was (and still remains) noticed in an extraordinary number of situations. Every business school student learns that 80% of a manufacturer’s sales (and complaints!) comes from 20% of customers. Some readers are familiar with power laws like Zipf’s law about the population of cities (the largest is observed to be twice the size of the second largest, and that second one is twice the size of the third largest and so forth), the prevalence of particular words in the written work of virtually all known languages, and even observations about the relative citations of academic articles. In these examples, the distributions are right skewed, and thus even steeper at the tail than would be expected from a merely exponential function.<sup>17</sup>

There is a substantial and fascinating literature on power laws, but for present purposes the thing to see is fairly intuitive. It is that regressions, or least-squared methods quite generally, will produce serious mistakes when

---

15. Thus, if a worker’s salary increases by 3% a year, the dollar increase in year 10 will be much greater than that in year 2, but the lifetime income pattern can be formulated as a straight line because of the regular 3% change. Linear regressions tend to be more useful than non-linear curves for reasons that can be intuited but that are not critical here.

16. In summary, minimizing least squares, when working with linear data under certain assumptions, provides the best linear unbiased estimates. See Quinlan Lee, *OLS, BLUE and the Gauss Markov Theorem*, U. WATERLOO ECON. SOC’Y (Mar. 1, 2017), <http://uweconsoc.com/ols-blue-and-the-gauss-markov-theorem> [https://perma.cc/4V3G-KHGT].

17. For these and many other examples of power laws, see William J. Reed, *The Pareto, Zipf and Other Power Laws*, 74 *ECON. LETTERS* 15 (2001); and PER BAK, *HOW NATURE WORKS: THE SCIENCE OF SELF-ORGANIZED CRITICALITY* (1997).

the tail of a distribution is steeper than expected. Errors will begin to fall farther and farther from one side of the line fitted to the expected distribution, and the model will experience high heteroscedasticity. One way to think about this is that if the empiricist is fitting a function to data with an expectation of some exponential function at the tail, least squares—which is the backbone of regression techniques and thus of empirical work in law and economics—will do an increasingly poor job when that tail is really much steeper than what was anticipated by observations marking an earlier shape of the distribution. The errors will be greater as we move along, and each error (whether or not squared) adds bias to the fitted curve.

A yet easier way to think of this is that if only the regression technique accounted for the presence of a power-law function, it could do a much better job. But how do we, or the statistical packages used by beginners,<sup>18</sup> know when we are facing a power law or other inconvenient function? The answer is that a *theory* is required. Indeed, the power-law literature is full of attempts to explain observed power laws. Some look for evolutionary explanations<sup>19</sup> and some to random-walk theory<sup>20</sup> (which can present another problem for empiricists), but this is obviously not the place to convince the reader to share my fascination with power-law distributions. Instead, the point is that it is more important to understand why power-law

---

18. Stata, SAS, R, and other regression software often require (depending on the type of regression) additional coding packages which are user-written and circulated online.

19. A famous example, Kleiber's Law, demonstrates both the explanatory significance of power laws and the difficulty academics face when they identify such a law. Max Kleiber found that animal metabolism, long known to be correlated with body size, scales at  $3/4$  power of the animal's body size, contradicting the previous wisdom that metabolism scales at a  $2/3$  power, owing to the relationship between surface area and volume (the way a 3x3x3 foot box contains 27 cubic feet). Kleiber's law has since been refuted and reinvigorated countless times. See Karl J. Niklas & Ulrich Kutschera, *Kleiber's Law: How the Fire of Life Ignited Debate, Fueled Theory, and Neglected Plants as Model Organisms*, 10 PLANT SIGNALING & BEHAV. 7 (2015). There is surely *some* relationship between metabolic rates and body size (or surface area or perhaps a combination of the two), as theory suggests, but it is not clear that evolutionary pressure leads to a universal relationship of the kind Kleiber announced. Moreover, at some point, the observations motivated by Kleiber's Law, often suggesting log-log transformations where power-law distributions are to be expected, smack of some retrofitting. After all, if we first try a linear regression, then a log transformation, and then log-log transformations, eventually something will fit the available data with some reasonable level of confidence, or so it seems. This appears to be what transpired, as new data about previously ignored species toppled the previously announced law, or suggested another. In any event, even after all these years, Kleiber's Law and its "exceptions" continue to defy straightforward theories. Early observations suggested a particular distribution and helped develop a convincing theory, and then further data challenged that theory but has not quite produced an improved theory. As is currently the case in law and economics, young empiricists were complimented for the acquisition and development of new data sets, but the quest for a comprehensive theory regarding this relatively small question has not succeeded. The Kleiber's Law industry also presents an interesting counterexample to the familiar claim that "it takes a theory to beat a theory."

20. A well-known starting point is PAUL H. COOTNER, THE RANDOM CHARACTER OF STOCK MARKET PRICES (1964). Random growth theory has also been applied to the hot topic of growing income inequality at the right tail, though other theories, like super-star returns, currently seem ascendant. Xavier Gabaix et al., *The Dynamics of Inequality*, 84 ECONOMETRICA 2071 (2016).

(or other) distributions arise, in order to understand where else they might be found. Without such theories, common regression techniques are misleading and the empiricist's conclusions will soon fail to replicate. In short, regression techniques (with recognizable distributions) form the backbone of empirical work in law and economics, and these regressions are often unjustified—and likely lead to misleading results.

### II. The Mistaken Assumption of Single Distributions

A second important reason to be skeptical of regressions, and especially so in law and economics, is that true distributions may change course over time or over another variable. Instead of fitting a curve over all available data (even after strategically and perhaps unwisely excluding some outliers<sup>21</sup>), so that many observations are shoved into one peg, in reality these observations are likely to fit one pattern for one period of time. Then, after some intervening event, the observations might fit another pattern, depending perhaps on another identifiable variable (such as a different time period). A theory is required in order to know whether to fit a curve to all the available data, or rather to fit a curve to an identifiable subset of the data, and then to expect a different distribution for the remainder. In the process, the theory will identify the transition point at which we need to begin the new distribution. This is easy to imagine if some, but only some of the data, follows a power law. A curve may simultaneously fit the relatively flat tail (only the first part of the data) very well, but then be a poor fit for the spike-end of the distribution. In artificial intelligence and other fields, this is sometimes referred to as a “transition,” meaning that beginning at some point, or subject to some variable, the data can be divided, or split, in two or more parts, so that different distributions best fit the several parts.

Note that this is more than a turning or inflection point. In the absence of a theory or two, the empiricist will not know how to divide the treatment and control groups, and unbalanced groups will be created. For example, Pareto's original intuition about wealth distribution does not offer an excellent fit for subsequent data about wealth, and it has been observed that

---

21. Omitting outliers haphazardly can be disastrous, especially when the assumed underlying distribution (typically a normal distribution for linear regressions) does not match up with reality. See, e.g., Carlos Fajardo et al., *One Needs to Be Careful When Dismissing Outliers: A Realistic Example*, U. TEX. EL PASO DEP'T OF COMPUTER SCI. (2016), [https://scholarworks.utep.edu/cgi/viewcontent.cgi?article=2044&context=cs\\_techrep](https://scholarworks.utep.edu/cgi/viewcontent.cgi?article=2044&context=cs_techrep) [<https://perma.cc/JYS2-DY6V>] (providing an example of a standard technique for omitting outliers that destroys the value of available data); see also Harvey J. Motulsky & Ronald E. Brown, *Detecting Outliers When Fitting Data with Nonlinear Regression—a New Method Based on Robust Nonlinear Regression and the False Discovery Rate*, 7 BMC BIOINFORMATICS 123 (2006) (explaining the difficulties of non-linear outlier omission and providing a possible solution). The empiricist can try to rely on a method announced before observing the data set, can instead use personal judgment, or can try to partner with a theory about the available data and issue.

if one applies an exponential distribution to a majority of the population, and then a power-law distribution just to the right tail, the overall fit is superior.<sup>22</sup> Two distributions are better than one, so to speak, and theorists have worked on theories that explain or suggest this bifurcation. In law, we might imagine that the distribution of data will follow one curve in some states and a very different curve in others. Alternatively, data might follow one curve before a certain date and another after it. Similarly, results might follow a different distribution for serious crimes than for minor transgressions. Empiricists are quick to think of this claim as one calling for holding the omitted variable constant, but this is not enough. The very distribution of the data might be different on one side of the excluded variable than the other. By different, I do not mean that the data is bimodal or multimodal, for that describes situations in which there is a single distribution with several peaks. The problem instead is that data might, to repeat the earlier example, follow a power-law distribution after a given event, or depending on one or more variables, so that we need to look for (at least) two distributions: one “before” the event and one following it. But, we do not know the dividing line without theory.<sup>23</sup> Some data is not necessarily better than no data if the available data is inspected for the wrong thing.

Another way to think about this is to ask about the nature of omitted variable problems. An empiricist might want to study whether increasing expected prison sentences brings about a decrease in crime. Imagine that the empiricist shows a strong relationship between threatened sentences and crime rates. The work may have taken advantage of a natural experiment in which the legislature instituted more severe sentences that were widely advertised. The implication is that, subject to various costs and benefits, it might (or might not) be worthwhile to increase penalties for convicted criminals. When such a finding is presented at an economics workshop, it is inevitable that most of the hands raised point to omitted variables that might be critical. Perhaps there was more investment in police forces over time, and this is what discouraged criminal activity. Maybe some prisons offered better job training or other conditions that made a prison sentence less of a deterrent. Perhaps juries responded to the advertised increase in required sentences by being more likely to find defendants innocent.

The empiricist will have foreseen many of these objections and will have tried to test their influence with various regressions. In some

---

22. See Thomas Lux, *Emergent Statistical Wealth Distributions in Simple Monetary Exchange Models: A Critical Review*, in *ECONOPHYSICS OF WEALTH DISTRIBUTIONS* 51 (A. Chatterjee et al., eds., 2005).

23. To avoid retrofitting, there might be cases (like wealth distribution) where two distributions, and an apparent transition, seem to work well. With some imagination a theory can be developed, and then tested on future data about wealth distributions. Again, the point here is not to say that data is useless, but rather that data and theory working together, sometimes in one order and sometimes in the other, is likely to be an important part of the future of law and economics.

disciplines this is easier than in others, because there are many observations. Omitted variables cannot simply be tested on their own, because there are likely to be joint and confounding effects among them. Variable *A* might be important only when variable *B* is also present, and variable *C* might be offset by the presence of variable *B*, and so forth. The fewer the observations, the more difficult it is to test these effects. Empiricists are likely to compensate for this shortage of observations, and even for the absence of any observations, with certain combinations of variables, by interpolations—but this requires some assumptions about underlying distributions.

To be sure, extrapolations and interpolations are reasonable when there are straight lines to be imagined on the basis of theories. But with zig-zags or other omitted variables that completely change the underlying distribution on one side of the line to be filled in by interpolation, the process is misleading or simply impossible. This is especially so when the empirical work is said to have implications for further increases in prison sentences. There are many reasons to think that what is true about behavior when a sentence is increased from three years to five, and then from five to nine, will tell us little about what to expect for an increase from nine to twelve years. Indeed, even if observations about increasing sentences from three to five, and then from nine to twelve years, fit one familiar distribution, there is great room for error with a nonlinear regression when attempting to say something about an intermediate increase from five to nine years, and especially so when this increase in sentence duration (and associated criminal behavior) is to take place in years following the studied observations.<sup>24</sup> Unfortunately, it is common for empiricists to lump the entire available sample, often for lack of a theory; this lumping is prone to reversal, a topic discussed in Part III. And if the empiricist tries to account for years of prison in this example, it is likely that unobserved counterfactuals come into play.

The larger point here is that the testing of omitted variables requires one to identify, or even imagine, these variables, and thus to have a theory of how the world works. Moreover, any test of the importance of a given omitted variable requires the empiricist to have a theory about the underlying (real) distribution of data, if we had a full set of observations. The latter problem can sometimes be handled with very large data sets—but these are normally unavailable where law is concerned. Part III now turns to another problem associated with limited data sets. The point will again be that a good theory is required before saying much about a data set.

---

24. Thus, if a single curve fits many increases in sentences over many years, at some point it would be reasonable to think it unlikely that the increase in penalty following a change from five to nine years would surprise us.

## III. Limited Data and Reversals

Most techniques for dealing with omitted variables require significant amounts of data. Various paths among variables and outcomes need to be examined.<sup>25</sup> For example, in order to judge the importance of, or causal relationship between, cigarette advertising and cancer, the empiricist needs significant data to evaluate the degree to which advertising brings about (or can discourage) smoking, and then the degree to which smoking rather than genetic or environmental variables, or various combinations of these variables, does the work. Serious scientists took many years to evaluate these connections and to be sure that cigarette advertising (and even smoking) was harmful. They did this with very large data sets, but without easy access to controlled experiments. In law and economics, we have somewhat better opportunities for natural or controlled experiments, but, unfortunately, we have at least as many variables to confront, and much smaller data sets.

As a co-author and I have emphasized,<sup>26</sup> relatively large data sets allow for the division of data and, among other things, some circling around omitted data concerns. When separated segments are small, any findings are unlikely to be statistically significant. An interesting subset of this omitted variable problem concerns reversal paradoxes, where the empiricist can fool us (and herself) into believing (not just the wrong strength of an effect but even) the wrong direction of a variable's impact.

Imagine a firm looking to make an offer to one of two summer associates, Kim and Kit.<sup>27</sup> The firm decides to score the associates on assignments given to them while they summer at the firm. During the first month of the summer, Kit is given 1 extensive corporate assignment and deemed to have done a poor job on it. Meanwhile, Kim is given 4 assignments in that department and is graded as a success on 1 of them. During the second month, the two are assigned to the environmental group. Kit receives 5 assignments and succeeds on 4. Kim is judged to have successfully completed the 2 assignments given in the same department. The firm tabulates the reviews and decides to hire Kit because Kit impresses on 4 of the 6 assignments, while Kim impressed partners on just 3 of the 6. But then a partner points out that perhaps the corporate assignments were simply more difficult than the tasks judged by the environmental group (difficulty is, after all, an omitted variable). Indeed, Kim performed better than Kit on the corporate assignments, and *also* better than Kit on the environmental projects. Each department would prefer Kim over Kit, even though Kit's overall score was superior. This is a classic reversal paradox. In this

---

25. A surprisingly simple yet helpful way to conceptualize this is with causal models used in combination with regression techniques. See PEARL & MACKENZIE, *supra* note 1, at 135-65.

26. Fagan & Levmore, *supra* note 5.

27. The example here extends one constructed in Fagan & Levmore, *supra* note 5, at 26-28.

## The Eventual Decline of Empirical Law and Economics

example, the reversal easily came about because the summer associates were not (and probably could not be) given the same assignments, and not even the same number of assignments in each department. Note that a further, or double, reversal paradox is possible, once we allow for other omitted variables. Perhaps partners tend to give high scores in the morning, and Kit was always evaluated in the afternoon. Had both candidates for employment been evaluated in the morning, it might have been Kit who would have been thought superior in the morning and *also* in the afternoon, and indeed by both departments.

There is no end to the omitted variable problem, and its capacity to change results. Even if Kim and Kit had been given the same assignments, from the same partner and at the same time of day, reversal would continue to be a danger because of the candidates' dissimilar innate abilities and characteristics, as well as the mix of cases assigned to them, and the kind of work the firm expects to have in the future. Kim may be bilingual and able to complete immigration cases and cross-border tax cases more quickly than Kit. But Kit may be better at managing work-life balance and coping with stress than is Kim. Any model that omits variables that account for those dissimilarities is fragile.

Again, it is common to question researchers about omitted variables. In the little example offered here, designed to highlight the fragility of models based on a small number of observations, when Kim is preferred after the initial analysis of performance, someone whose intuition was to favor Kit might have pointed out that Kit's assignments were more difficult or that Kit was evaluated by partners who tended to be tough graders. If the omitted variable were properly included, the result would have been different. But in most cases this would mean that one department would favor Kit over the declared winner, Kim. The overall scores might be different, with some disputes about how to weigh the factors that contributed to these scores. The remarkable thing about the special case of a reversal (or Simpson's) paradox is that an omitted variable causes *both* departments to favor one result, while the overall, combined score still favors the opposite result, even when all known variables are included. The practical and often startling lesson is that even when an empirical study is questioned because of some omitted variable, and that variable is included in further study, the result may still favor *X* over *Y*, even though *Y* is superior to *X* in settings for which data is available. This will not occur if one knows exactly how to weigh one or more omitted variables, but the weight itself is often unknown or unmeasurable and can be thought of as an additional omitted variable. Assigning a weight or properly modeling a relationship is often difficult, and it is hard or even impossible to know when it has been done correctly.

Reversal paradoxes should be understood as a subset of the problem of omitted variables, but it is a particularly interesting subset both because of the startling reversals and because these problems are more difficult to

solve than other omitted variable problems that can be decoded with additional testing and clever study design. In some cases the larger set of data is what matters, and in other cases the divided data ought to carry the day. In the example just offered, I think readers would have agreed that Kim ought to be hired, once it was shown that both departments favored Kim; agreement is less likely after the time-of-day variable is introduced. Dividing and validating data is a practice that avoids many problems in empirical work, and it can reduce the risk of reversals. If there are 20,000 patients with a disease, and a scientist wants to test combinations of drugs, it is usually wise to find the winning combination on a group of 10,000 randomly chosen patients, and then test this finding by applying it to the previously untested, or set-aside, 10,000. In the necessarily smaller case, if Kim and Kit are evaluated over multiple summers and some term-time opportunities to work for the firm, and Kit earns higher total scores, it is less likely that Kim was unknowingly given more difficult assignments or graders, or was evaluated at an unlucky time of day, enough to reverse the result reached by taking total scores.<sup>28</sup> A hidden reversal paradox is less likely as the number of comparisons increases. Note that what we think of as a conventional division (here, by law firm department or by summers, or both) is hardly random, and there are many ways to divide the candidates' performances. Data division, followed by validation, is almost always a good and workable idea when big data are available,<sup>29</sup> but it is critical to have a random division of the data. This has become a best practice in modern data analysis, but it has not yet come to law and economics. The larger and more obvious point is that law and economics empiricism is especially likely to be unpleasantly surprised by reversal paradoxes because it often makes use of relatively small data sets. The problem is acute when non-linear regressions are in play.

#### IV. Theory-Data Interdependence

I have emphasized that the possibility of inconvenient distributions and multiple and unforeseen omitted variables should cast doubt on many conclusions normally attributed to empirical work. This Essay, like some work in artificial intelligence, thus leads to the idea that data and theory need to work together.<sup>30</sup> Data can show associations but not causation, unless the empiricist knows what to inspect. Even then, the empiricist faces

---

28. Similarly, a track coach may seek the fastest runner for the team by averaging times in a given race. One runner may have the lowest average time, but another may win when the running path includes hills, when the weather is cool, when the race is run in the morning, and when the race is run indoors. But there are other ways to divide the races, and these hidden variables can bring about reversals. *But see infra* note 29.

29. Predictive models are generally trained and tested with set-aside data. Current techniques, such as k-fold cross-validation, split the data into a number of random partitions for estimating model performance; accuracy rates are assessed for each partition, or fold, to determine adjustments to the model.

30. See PEARL & MACKENZIE, *supra* note 1, at 349-53.

an uphill climb. At the same time, the observation about data-theory interdependence raises the question of where theories come from, and also points to the importance of asking whether a theory precedes or follows data investigation.

In some cases a theory is so convincing that it can survive the difficulties of confirmatory data. A counterintuitive theory is probably more valuable than a counterintuitive and one-time empirical finding. Consider, for example, the idea of “winner’s curse.” In a standard and familiar English auction, the winner is the one who submits the highest bid, and only the winner pays (the amount bid). It is likely that with many bidders, the median bidder’s valuation is a good estimate of the market value of the item up for auction.<sup>31</sup> As it turns out, this “wisdom of crowds” claim can be demonstrated in a class to amazed or amused students with such regularity that we might think of it as empirically verified by playing game after game (with real money in order to avoid claims that fun games do not cast light on the real world). In any event, if the median bidder is likely to arrive at something close to the correct valuation, it is apparent that the high bidder, or auction “winner,” is likely to overvalue the item and bid up the price to the point of overpayment.

More serious empirical evidence, away from a class of inexperienced bidders with mixed motives, is difficult to obtain. For one thing, seasoned participants will learn to adjust for the winner’s curse; one who values an item at twenty, might learn to bid only up to fifteen, depending on his experience in prior auctions or, if mathematically inclined, on his assessment of the number of bidders and the distribution of expected values. Over time, then, the winner’s curse may disappear, or potential bidders may learn from their mistakes and avoid auctions, up to the point where the seller no longer uses auctions. The winner’s curse logic is so strong that empirical evidence is probably unnecessary. On the other hand, it is only fair to note that the theoretical insight came about because of observations in the field.<sup>32</sup>

I feel obliged to point out that winner’s curse as applied to reported results is yet another reason to be skeptical of empirical claims. If empiricists are confident of a result when it meets some threshold, like a ninety-

---

31. Richard H. Thaler, *Anomalies: The Winner’s Curse*, 2 J. ECON. PERSP. 191 (1988). Winner’s curse is an offshoot of the Condorcet Jury Theorem, an idea that may also be obviously correct, once thought through, such that empirical verification is probably unnecessary. David Austen-Smith & Jeffrey S. Banks, *Information Aggregation, Rationality, and the Condorcet Jury Theorem*, 90 AM. POL. SCI. REV. 34, 34-35 (1996).

32. On the history of winner’s curse and empirical observations leading to its formation and then testing, see R. Preston McAfee & John McMillan, *Auctions and Bidding*, 25 J. ECON. LIT. 699 (1987). When a passing observation inspires thoughtful analysis and then a new theory, as it did in the case of winner’s curse, we might say that a kind of bad empirical work (with no controls, a small data set, and no division of data) brought about a theory (which turned out to be difficult to test). This seems quite common, but it is an advertisement for intellectual curiosity rather than for serious empirical work.

five percent chance that the connection it finds is not the product of randomness, then it stands to reason that the finding overstates the likely existence of a non-random relationship. Along with the other problems noted earlier in this Essay, it is easy to see why empirical law and economics is likely to face a replication crisis. Some attempts at replication will fail because of previously omitted variables, some because of unrecognized power-law (or other) distributions, and some because only a subset of results has been reported.<sup>33</sup> In the latter case, the replication crisis may seem less serious, inasmuch as an effect is correctly identified, albeit misestimated. But where costs and benefits are concerned, the strength of an effect is important and misestimation may as well be considered part of a replication problem. This can be thought of as a form of regression to the mean; the advertised claim is likely to be an outlier. It is more meaningful when the empiricist expresses confidence that  $x > y$  or that  $x$  is associated with  $y$ , than when the empiricist says that every additional unit of  $v$  is likely to bring about some number of units of  $w$ . The problem described here is of course exacerbated by the familiar complaint about the difficulty of publishing papers showing statistically insignificant results.<sup>34</sup> The median paper is not the median statistical study. Moreover, it is difficult to adjust correctly for this bias.

It is commonly believed that a theory, or hypothesis, has little value unless it can be tested, but testing is often not easy. This reality devalues empirical work at least as much as it casts doubt on theoretical claims. A positive prediction is subject to the objection that some other variable might explain an observed result. Similarly, a negative result does not prove much if the proposed test is itself confounded by an unexpected omitted variable. Thus, the theorist is frequently asked, “What would prove you wrong?” and yet the unwanted result may not prove the theory wrong if it is the product of an unanticipated variable.

To be sure, however imperfect empirical work may be, positive results usually and correctly add to the likelihood that the theory is correct. We can think of this as an example of Bayesian updating. If, for example, the theorist claims that higher taxes will lead to less investment in a jurisdiction, and this is indeed observed, the skeptic might name a hundred other reasons why investment declined. And yet, if multiple tests or natural experiments repeatedly show reduced investment, it is sensible to think that the theory about the impact of higher taxes might indeed be correct. Causation has hardly been demonstrated to a thinking person’s satisfaction, but even the most skeptical observer will be less likely to wager that

---

33. The need for replication might suggest an increase in demand for empirical work, and along with decreasing costs of production, this might suggest that empirical work will increase rather than decline in relative terms, as suggested here. On the other hand, it is rare for demand to increase for things when they are shown to be of lower value than previously imagined.

34. Ana Mlinarić et al., *Dealing with the Positive Publication Bias: Why You Should Really Publish Your Negative Results*, 27 *BIOCHEMIA MEDICA* 3 (2017).

investment will not decline the next time taxes are increased, even in previously untested jurisdictions.

It is plain that a theory can be made more attractive with supporting data, despite the skeptic's objection that a regression result suffers from some omitted variables. At the same time, a theory is often less easily rejected in the face of conflicting data. This is because an attractive theory changes our perceptions, or priors, on its own. In any event, when theory and data evolve together, our understanding of phenomena is improved. Each one alone does some work, but data alone is unconvincing, while theory alone is, I believe, sometimes convincing – and especially so if it is difficult to think of a workable experiment or regression analysis that will come close to bringing matters to rest. The most we can say is that theories can be enhanced or disfavored because of empirical work, while data can to a more limited degree suggest new theories that might be tested. In terms of the evolution of law and economics, it is no accident that a generation of theorists was followed by a generation of empiricists. But in similar fashion, a generation dominated by empiricists is soon likely to be followed by a re-emergence of theorists. Ironically, this too is a theory that is not easily tested.<sup>35</sup>

It may be useful to offer a simple example, chosen almost at random, from practical matters that law and economics would like to influence, in order to demonstrate its practical importance. A defendant facing a serious criminal charge would like to know whether to choose a bench trial or take advantage of the right to a jury trial; similarly, in many civil cases there is an opportunity to insist on a jury trial.<sup>36</sup> A typical empirical study would show conviction (or civil liability) rates in various states, and try to say something about when defendants should prefer bench trials. Conventional wisdom focuses on cases where a legal technicality might favor the defendant and suggests that it is in these cases that a bench trial is to be preferred. In civil cases, a sophisticated approach might look at the wealth of the defendant, the results in previous cases assigned to a given judge, and so forth. There is a large body of data, compared with many other decisions facing a litigant, but this volume quickly declines once the empiricist (or artificial intelligence), aiming to learn about the impact of a jury trial, sets to work on the relevance of a particular factor, including the state in which the trial is to take place; the gender of the defendant, plaintiff, or judge; and the apparent social status of these parties, not to mention their

---

35. One objection to this claim is that empirical methods have improved greatly over the last generation (an undefined time period here), but it is hard to say the same thing about theories. The common view that theorists in law and economics have plucked the low hanging fruit, and this makes it more difficult for newcomers is, however, offset by the tendency of new empiricists to over-claim in the interest of making a splash or getting accepted by law journals.

36. See, e.g., Nancy J. King et al., *When Process Affects Punishment: Differences in Sentences After Guilty Plea, Bench Trial, and Jury Trial in Five Guidelines States*, 105 COLUM. L. REV. 959 (2005).

choices of what to wear or how loudly to speak in court. Various empirical studies also claim that the time of day matters, as might the political persuasion of the judge, the size of the jury, the selection of jury members, and so forth.<sup>37</sup> The list of plausible variables is astonishingly great, and we must take into account interactions among these variables. In addition, both the jury and opposing counsel might learn something from whether a party chooses to forego the right to a jury; it is easy to see that a party might in some situations benefit by demonstrating that it has chosen or rejected a jury trial with the flip of a coin. In a sense, the party is trying to correct for what might be seen as selection bias.

It is noteworthy that what is typically at stake here is the question of whether to settle a case or to accept a plea bargain, and these decisions are not all-or-nothing, but surely depend on the size of the offer that is made. On the one hand, this need to understand the scale of an effect suggests why data alone is unlikely to get us very far. Moreover, an empiricist who aims not to help one side or the other in these cases, but to improve the legal system, faces an even greater problem. This empiricist wants to know when juries are likely to reach the right result. In turn, this requires some knowledge of whether a defendant was indeed guilty or negligent. This information is hard to obtain and makes the entire venture yet more difficult. It is tempting to say that difficult questions about medicine are, for the empiricist, far easier than the most straightforward binary questions in law. Empiricists in law and economics seem likely to study these variables and produce many doctoral degrees in the process, but it is doubtful that much that is privately or socially useful will be learned other than a conclusion that either overclaims or concludes that “further study is needed.”

At the same time, this typical, or apparently straightforward, sort of case (involving the decision to choose a jury trial), beginning with a relatively large data set, does not support the central claim of this Essay. I aimed to show that theory and data need to work together, but the example in this Part is no more supportive of theory than it has been of data. Theories, or hypotheses, about the impact of juries are easy to manufacture — though an “obvious” theory alone is unlikely to produce a doctoral degree or a tenured position. No theorist gets much credit for a claim about why a requirement of unanimity might be better for the defendant. And no theorist on her own gets any credit for opining that judges are more likely to make certain decisions before or after lunch. It is too easy to state these hypotheses before or after any empirical evidence, and indeed the first claim is unlikely to get attention even if it comes with supporting empirical evidence. Even a more surprising result can be shrugged off; if it could be demonstrated that judge or jury decisions were influenced by time of day

---

37. Unsurprisingly, these results get attention and are then often found not to replicate, likely because of an omitted variable. With regard to the time of day claim about judging, see Keren Weinshall-Margel & John Shapard, *Overlooked Factors in the Analysis of Parole Decisions*, 108 PROC. NAT'L ACAD. SCI. U.S. E833 (2011).

## The Eventual Decline of Empirical Law and Economics

or gender, for instance, once judges learned of these claimed biases it is likely that there would be some adjustment. As in the case of winner's curse, or many theories (and empirical findings) in finance, the real test of a result is whether it applies to set-aside, and in many cases to future, data. The combination of all these difficulties suggests that an interesting (non-obvious) theory followed by the absence of repeatedly contrasting data is an imperfect but promising way to think about the co-evolution of theory and data in our quest to better understand law or to change it for the better.

Note that this Essay is not claiming that all questions are similar to the question of jury choice. The jury example was chosen because it offers a much larger set of data than is usually available to law and economics scholars, and large data sets are often needed to overcome the problems posed by omitted variables. I continue to have faith in theories like winner's curse, even though large data sets might prove to be tarnished, as discussed earlier. The value of such theories comes from their surprise and plausibility.

### Conclusion

The aim here has not been to denigrate regressions, but rather to make the case for the co-evolution, and even co-determination, of theoretical and empirical insights. Law and economics has come to be dominated by regression analysis, and new entrants to the field more often than not make their reputations with this subset of empirical work. Most of these projects make questionable assumptions. When empiricists are challenged about the assumptions implicit in their models, they often respond that they are simply following the accepted practices in the field. In my own experience, to question these models and the conclusions they suggest is to be labeled as one who is hostile to empirical work and its ascendance. This is unfortunate and unscientific.

The problems with empirical law and economics do not make a case for unchallenged theoretical work. Weaknesses here do not make for victory there, in the realm of theory. From the beginning, theorists were challenged about their assumptions, even by generalists who needed no special training in economics, but who are ill-equipped to challenge empirical work on its terms. Empirically minded scholars insisted that a theory requires empirical evidence, usually in the form of a test that could be undertaken and that could demonstrate that the theory was false. Most of the arguments in this Essay have been directed at empirical work, but these challenges also mean that theories are weaker than might have been imagined because they cannot easily be shown to be true or false. Theories can be imperfectly tested, and over time they gain or lose support. None of the problems with empirical work makes the co-evolution of, or teamwork between, empirical and theoretical work magically immune from the challenges put forward here. It is more realistic to think that our confidence in

theories as well as in empirical work will grow when the two work in partnership and point in the same direction.